

Puntuaciones relacionadas con las normas

Luisa Lezama

luisalezama@yahoo.com

Escuela de Psicología-Universidad Central de Venezuela

Resumen

La interpretación de las calificaciones de un instrumento depende de su tipo. En el caso de las pruebas normativas, el constructor podrá recurrir a puntuaciones lineales o normalizadas, dependiendo de la forma de la distribución de las calificaciones directas, y cuando la variable examinada cambie sistemáticamente, podrá recurrir a las escalas de desarrollo. Los datos normativos se pueden presentar en cuadros de conversión simple, múltiple o de perfiles, y se expresan mediante normas nacionales, regionales o locales que permitirán hacer uso de la equiparación de puntuaciones, a la hora de efectuar comparaciones intra e interpruebas. Este artículo caracteriza cada una de estos contenidos y especifica los procesamientos estadísticos que es necesario realizar para preparar las normas de una prueba en cualquiera de sus versiones.

Palabras clave: calificaciones estándares lineales, calificaciones estándares normalizadas, escalas de desarrollo

Recibido: 21 de febrero de 2011

Aprobado: 18 de marzo de 2011

SCORING RELATED TO NORMS

Abstract

The interpretation of the instruments scoring depends on their type. In normative tests, the evaluator can use lineal scores or normative ones depending on the distribution of direct grades, and when the examined variable changes systemayically, he will be able to apply developmental scales. Normative data can be presented by means of simple, multiple or profiles simple conversion tables. These data are expressed through national, regional or local norms which will permit the use of even scorings when comparing inter and intra tests. This article characterizes each of these contents and specify the statistical procedures needed to prepare the norms of a test in every version.

Key words: standarized lineal scores, standarized normed scores, developmental scales

Se afirma que un test psicológico es una medida objetiva y *estandarizada* de una muestra de conducta; el término estandarizada es de interés central acá ya que conlleva que los puntajes de las personas en un instrumento provienen de la aplicación de reglas bien desarrolladas. Dichas reglas exigen que a todos esos sujetos les sean administrados los mismos reactivos bajo condiciones idénticas; que la corrección del instrumento se efectúe siguiendo el mismo protocolo; y que la valoración del puntaje se haga en función de cierto tipo de calificaciones vinculadas directamente con el tipo de prueba que se trate; es decir, si esa prueba es criterial se emplearán, por ejemplo, calificaciones de pase, escalas de clasificación, porcentajes de respuesta, pero si el instrumento es normativo se recurrirá a las llamadas calificaciones *relacionadas con las normas*. Estas últimas representan el objetivo central de este documento.

Las calificaciones relacionadas con las normas son aquellas que se emplean cuando el interés central es comparar el puntaje de una persona con el puntaje de un grupo de personas con quienes comparte características, y al que se denomina *grupo normativo*. El grupo normativo provee el marco de referencia con relación al cual se realiza la interpretación de un puntaje particular, por lo cual debe ser cuidadosamente descrito en el manual de la prueba, dando evidencias de que dicho grupo es no sólo numéricamente adecuado, sino cualitativamente representativo de la población a la cual se dirige el instrumento. El usuario de prueba deberá examinar esa descripción minuciosamente, con el fin de verificar que su examinado puede ser enmarcado dentro de esas características, con lo cual, lo que es cierto para ese grupo normativo, también lo será para su evaluado.

Una vez que un especialista aplica una prueba debe proceder a corregirla, obteniendo para cada examinado lo que ha sido etiquetado en la literatura como *puntuación bruta, directa u original*. En los tests de capacidad y rendimiento esas puntuaciones directas pueden ser el número de respuestas correctas, el número de errores, el tiempo requerido para contestar la prueba, etcétera; en los instrumentos que miden personalidad, las puntuaciones brutas suelen ser las opciones predominantes en un grupo de referencia o las opciones de una escala Likert.

Dichas puntuaciones brutas, directas u originales carecen de significado por sí solas y adquieren sentido sólo cuando se les compara con la información del grupo normativo provista en el manual del instrumento. Así, un puntaje bruto de 70 en un tests de razonamiento abstracto no significa nada hasta

que no se contraste con datos referenciales como, por ejemplo, la media del grupo; entonces, si la media de éste fuera 30 puntos, podría afirmarse que la persona que calificó con 70 salió muy bien respecto de ese grupo, pero si la media hubiese sido 100 se debería decir que obtuvo un bajo desempeño. En suma, un mismo puntaje bruto tiene implicaciones particulares que son dadas por las puntuaciones de referencia, regularmente etiquetadas bajo el nombre de *normas*.

Las normas se emplean cuando la medida que se toma con un instrumento tiene una importancia pragmática más que teórica, de modo que son esenciales en instrumentos con validez predictiva o con fines diagnósticos, pero no en los que poseen validez de constructo. Es menos probable que las medidas de constructo requieran de normas, ya que su interés principal se centra en la investigación y en la documentación de teorías y puede que nunca se usen para la clasificación o la selección individual, pero los instrumentos empleados para efectuar evaluaciones de personas particulares o para hacer estimaciones sí requerirán de un contenido validado y de normas construidas con toda la rigurosidad del caso.

TIPOS DE PUNTUACIONES TRANSFORMADAS

Las normas de un instrumento estandarizado de medición se conforman a partir de la transformación de puntajes directos, una vez administrado dicho instrumento a un grupo representativo de la población al que el mismo se dirige. La selección de la escala a la que se realizará la transformación de esas calificaciones originales, está sujeta a la forma de la distribución de dichos puntajes, la finura discriminativa que se requiera y la naturaleza de la variable medida.

Como se recuerda, los supuestos fundamentales de la medición en psicología postulan que toda variable se distribuye a lo largo de un continuo direccional (va de menor a mayor), que cada persona examinada en esa variable se ubica en algún lugar de tal continuo, y que, si se examinara la variable en toda la población se obtendrá una distribución normal (simétrica y mesokúrtica). Tales postulados permiten que las puntuaciones expresadas en una escala ordinal, que es el máximo nivel de medida alcanzado por las pruebas, puedan ser tratadas como si se expresaran en una escala de intervalos, posibilitando el uso de los estadísticos correspondientes a este nivel de medición y de aquellos que requieren de normalidad de la distribución.

Considerando lo anterior, cuando se diseña un instrumento de medición cuyo objetivo es analizar la puntuación de una persona en atención a otras con similares características, lo conducente es verificar si la forma de la distribución de los datos de ésta es efectivamente normal, para poder así hacer uso, como se dijo antes, de los recursos estadísticos que la curva normal provee.

La información acerca de si la distribución de los puntajes del instrumento que se está diseñando es normal se obtiene, ejecutando para ese conjunto de datos, la prueba de *Kolmogorov-Smirnov*; dicha prueba compara la función de distribución acumulada de los datos observados (los de la propia distribución) con la de una distribución normal teórica, midiendo la máxima distancia entre ambas curvas.

Cuadro 1

Prueba Kolmogorov-Smirnov para los datos de la EAR-1

		Total prueba
N		123
Parámetros normales (a, b)	Media	84,499
	Desviación	9,446
Diferencias más extremas	Absoluta	,085
	Positiva	,081
	Negativa	-,085
Z de Kolmogorov-Smirnov		,945
Sig. asintót. (bilateral)		,333

a: La distribución de contraste es normal

b: Se han calculado a partir de los datos

Como se aprecia en el cuadro 1, al ejecutar la prueba *Kolmogorov-Smirnov* para los datos de la Escala Agresividad Reactiva Forma 1 se obtuvo un $K-S(123) = ,945$, $p > ,05$; como el valor del estadístico es mayor que el valor crítico, se asume que la distribución obtenida es normal. Cuando la forma de la distribución de los puntajes brutos se ajusta a una curva normal, como en el caso del ejemplo, dichos puntajes deben ser convertidos a normas mediante *transformaciones lineales*, tal como se verá seguidamente.

Caso distinto ocurre cuando la forma de la distribución es “no normal”, como en la distribución de puntajes brutos de la Escala de Imagen Corporal. Como se aprecia en el cuadro 2, al ejecutar la prueba *Kolmogorov-Smirnov* para los datos de dicha escala se obtuvo que $K-S(201) = 3,347$, $p < ,05$; como el valor del estadístico es menor que el valor crítico, se rechaza la hipótesis nula, con lo cual se asume que la distribución obtenida es diferente a la normal teórica, es decir, la distribución es “no normal”.

Cuadro 2

Prueba Kolmogorov-Smirnov para los datos de la AR-1

		Total prueba
N		201
Parámetros normales (a, b)	Media	28,234
	Desviación	8,669
Diferencias más extremas	Absoluta	,237
	Positiva	,187
	Negativa	-,236
Z de Kolmogorov-Smirnov		3,347
Sig. asintót. (bilateral)		,000

a: La distribución de contraste es normal

b: Se han calculado a partir de los datos

En los casos en que la distribución de puntajes originales del instrumento en construcción sea “no normal” y ello se deba a problemas de muestreo, tal como se verá más adelante, dichos puntajes deberán ser normalizados a través de *transformaciones no lineales*.

Otro aspecto que se debe considerar cuando se va a escoger el tipo de puntuaciones normativas es la naturaleza de la variable examinada. Así, por ejemplo, si se conoce que la variable cambia sistemáticamente con el transcurso del tiempo, se pueden elaborar escalas de calificaciones que comparen la ejecución de un individuo con la persona promedio a varios niveles evolutivos; a estas escalas se les conoce como *escalas de desarrollo*, entre las cuales se pueden mencionar *los grados escolares*, *los cocientes intelectuales*, *las escalas ordinales*, etcétera, las cuales serán desarrolladas más adelante.

TRANSFORMACIONES LINEALES

Se recurre a este tipo de calificación cuando la forma de la distribución de los puntajes originales de la prueba asume una forma normal, es decir, es simétrica y mesokúrtica. Existen dos transformaciones que ofrecen esta posibilidad: *las puntuaciones típicas lineales directas y las derivadas*.

PUNTUACIONES TÍPICAS O ESTÁNDAR LINEALES DIRECTAS Z_x

Según Magnusson (1975), una transformación lineal implica *cambiar* la escala de manera que se modifique la media y/o la desviación estándar original, mientras se conserva idéntica la forma de la distribución y, en consecuencia, las posiciones relativas de los individuos en esa distribución transformada. Se cambia *la media* de la distribución original, por ejemplo, restando una constante a cada puntaje en la distribución, lo cual significa que toda la distribución se mueve a lo largo de la escala en la cantidad indicada por la constante. También se cambia *la desviación estándar* de la distribución original, sin afectar las posiciones relativas de los sujetos, por ejemplo, multiplicando por una constante, la cual da la relación entre la desviación típica original y la deseada.

Las transformaciones lineales son denominadas *calificaciones típicas o estándar* (z_x) y señalan la distancia, en términos de desviaciones típicas, a que se encuentra un puntaje original con respecto a la media de su distribución de puntajes originales. Las calificaciones z_x se disponen en una escala de intervalos iguales, con lo cual es posible realizar todas las operaciones matemáticas básicas, así como avanzados análisis estadísticos. Dichas calificaciones se obtienen por la fórmula
$$z = \frac{X - \bar{X}}{\sigma} = \frac{x}{\sigma}$$

En el siguiente grupo de datos, para el cual se calculó una media de 4,91 y desviación típica 1,86, al aplicar la ecuación anterior se puede determinar la puntuación típica z_x correspondiente a cada puntuación original en el test.

Cuadro 3

Puntuaciones típicas lineales directas

Puntuaciones en el test	Puntuaciones típicas (Z_x)
0	-2,6
1	-2,1
2	-1,6
3	-1
4	-0,5
5	0,1
6	0,6
7	1,1
8	1,7
9	2,2
10	2,7

Si se calcula la media y la desviación típica de tal distribución de puntuaciones estandarizadas z_x , se encontrará que éstas tienen media cero y desviación 1, por las razones que se exponen a continuación y que son tomadas de Ghiselli (1964):

a) Tiene media cero porque:

$$\bar{z} = \frac{\sum z}{n} = \frac{\sum \left(\frac{x}{\sigma}\right)}{n} = \frac{\left(\frac{1}{\sigma}\right)\sum x}{n} = \frac{1}{\sigma} \frac{\sum x}{n} \text{ pero como se sabe que } \frac{\sum x}{n} = 0 \Rightarrow \bar{z} = \frac{1}{\sigma} 0 = 0$$

b) La desviación estándar es igual a la unidad porque:

$$\begin{aligned} \sigma_z^2 &= \frac{\sum z^2}{n} \\ &= \frac{\sum (x/\sigma_x)^2}{n} = \frac{\sum (x^2/\sigma_x^2)}{n} = \frac{(1/\sigma_x^2)\sum x^2}{n} = \frac{1}{\sigma_x^2} \frac{\sum x^2}{n} \end{aligned}$$

Ya que $\frac{\sum x^2}{n} = \sigma_x^2$ se puede escribir que $\sigma_x^2 = \frac{1}{2} \sigma_x^2 = 1$

Dado que la desviación típica es la raíz cuadrada de la varianza, la desviación también será uno.

El rango de las puntuaciones estándar para la mayor parte de los grupos, no se extiende más allá de 3 desviaciones típicas por encima o por debajo de la media. El valor absoluto de una calificación típica indica la distancia a la que se encuentra la calificación original de la media de la distribución; su signo señala si esa calificación está por debajo (-) o por encima (+) de la media de la distribución. De modo que una persona que obtiene una calificación igual a la media de su grupo, recibirá una puntuación estándar de cero; la que puntúe por encima de la media obtendrá una calificación positiva; y la que se ubique por debajo de la media obtendrá un puntaje negativo.

Al transformar dos distribuciones de calificaciones originales diferentes a puntuaciones típicas, se convierten escalas diferentes a una misma escala, haciendo posible efectuar comparaciones interindividuales e intraindividuales. Si las distribuciones que se comparan tienen una forma normal, un mismo valor numérico de z_x en cada prueba tendrá exactamente el mismo significado, esto es, está a la misma distancia de su media respectiva.

PUNTUACIONES TÍPICAS O ESTÁNDARES LINEALES DERIVADAS

Las puntuaciones típicas o estándar lineales directas (Z_x) presentan un rango pequeño al oscilar entre ± 3 puntos, siendo allí donde radica su mayor inconveniente, ya que al hacer análisis se debe recurrir, por un lado, al uso de números negativos y, por otro lado, al empleo de decimales para expresar diferencias entre puntuaciones.

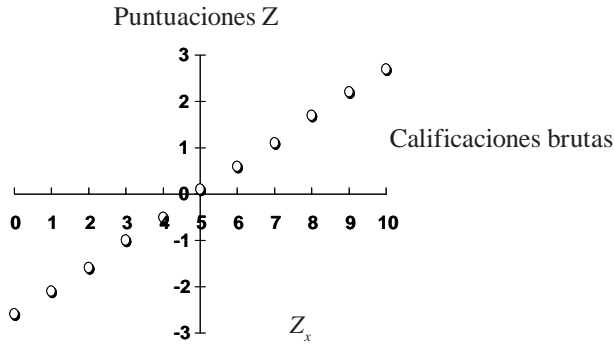


Figura 1. Calificación Z_x para cada puntuación original

Dado que, como se puede apreciar en la figura 1, al relacionar las calificaciones brutas (x) con las transformadas (z_x) se obtiene una relación lineal, es posible recurrir a la ecuación de regresión lineal simple

$$Y' = r_{xy} \frac{s_y}{s_x} (x - \bar{x}) + \bar{y} \quad \text{y así transformar las puntuaciones típicas lineales}$$

directas a otras escalas, que ofrezcan la ventaja de valores de media y desviación típica más grandes, permitiendo subsanar la limitación del empleo de números decimales y/o negativos, característicos de las escalas z_x . A estas calificaciones se les conoce con el nombre de *puntuaciones típicas o estándar lineales derivadas*.

Como la transformación lineal no implica cambio en la posición relativa de los sujetos, la correlación entre los puntajes originales y los transformados será de uno, con lo cual la expresión anterior se puede escribir como

$$Y' = \frac{s_y}{s_x} (x - \bar{x}) + \bar{y} \quad \text{que es lo mismo a escribir} \quad Y' = \bar{y} + (z_x)(s_y)$$

Donde Y' es una puntuación típica lineal derivada, Z_x es la puntuación típica lineal directa que le corresponde a un puntaje original x ; \bar{y} y Z_y son la media y la desviación típica de la distribución a la cual se hace la transformación.

La media y la desviación típica que se pueden emplear para expresar las puntuaciones estandarizadas derivadas son arbitrarias, pero la APA en 1974

recomendó el empleo de la escala *Z*, que ubica la media en 50 y la desviación típica en 10, como la usada por el *MMPI*. También suele usarse media 100 y desviación típica 15, como lo hace el *WAIS*; media 100 y desviación típica 16, como el Stanford Binet; media 500 y desviación 100, como *CEEB* y media 100, y desviación 20 de la escala *SAT*.

Si se toma como ejemplo a una persona que calificó con 7 puntos en una escala de media 4,91 y desviación típica 1,86, se puede observar que en una escala *SAT* de media 100 y desviación típica 20 le corresponde una

calificación de 122, ya que $SAT = \left(\frac{7 - 4,91}{1,86}\right) \cdot (20) + 100 = 122,4$.

Dicha calificación significa que una persona que calificó con 7 puntos en la distribución de puntajes originales le corresponde una puntuación típica lineal derivada específicamente en una escala *SAT* de 122, ubicándose a 1.1 desviaciones típicas por encima de la media de su grupo normativo.

Como se pudo apreciar en la ecuación anterior, es una forma general que incluye la transformación a puntuación típica lineal directa $\left(\frac{7 - 4,91}{1,86}\right)$ para luego multiplicar por la desviación y sumar la media de la escala derivada a la que se haga la transformación $Z_x (20) + 100$.

Cuadro 4

Puntuaciones típicas lineales derivadas a escala SAT y Z

Puntuaciones en el test	Típicas derivadas SAT	Típicas derivadas Z	
0	48	24	
1	58	29	
2	68	34	
3	80	40	<i>Puntuaciones directas</i>
4	90	45	$\bar{X} = 4,91$
5	102	51	$S = 1,86$
6	112	56	<i>Escala SAT</i>
7	122	61	$\bar{X} = 100$
8	134	67	$S = 20$
9	144	72	<i>Escala Z</i>
10	154	77	$\bar{X} = 50$
			$S = 10$

Las *puntuaciones típicas o estándar lineales derivadas* conservan las ventajas de las *puntuaciones típicas o estándar lineales directas* al indicar la posición relativa de la persona respecto a la media en término de desviaciones típicas, sin alterar la forma de la distribución original; además, permiten trabajar sólo con números positivos sin la necesidad de recurrir a diferenciar entre calificaciones a partir de decimales o a los números negativos, tal como se puede apreciar en el cuadro 4, donde se presenta la transformación a una escala *SAT* y a otra más común como lo es la escala *Z*, cuya media, como se dijo, es 50 y su desviación 10.

TRANSFORMACIONES NO LINEALES

Como se recuerda, las transformaciones no lineales se calculan cuando la forma de la distribución de los puntajes originales es no normal, en cuyo caso la transformación realizada cambia la forma de esa distribución, “obligándola” a comportarse como una curva normal. Entre las puntuaciones no lineales se pueden citar los percentiles, las calificaciones normalizadas directas y derivadas y las estaninas.

PERCENTILES

Esta transformación consiste en asignar a cada puntuación directa el porcentaje de personas que obtienen calificaciones inferiores a ella. De modo que informa el porcentaje de miembros del grupo normativo que obtuvieron calificaciones que igualan o superan un puntaje dado, así que mientras más bajo es el percentil más baja es la posición de la persona.

La unidad en que se expresa un percentil es en “porcentaje de personas” y no en la cantidad de algún rasgo psicológico, por lo cual un percentil no informa las diferencias de grado o cantidad en el rasgo particular, sino el número de personas con que un puntaje excede a otro.

Si se tienen los percentiles de tres personas $A = P_{70}$, $B = P_{60}$ y $C = P_{50}$, se puede decir que A es superior a B y éste a C , pero no se puede decir que la diferencia entre A y B en la propiedad que se está midiendo, es la misma en magnitud que la diferencia entre B y C , aunque sea la misma (10) en puntos percentilares.

Se observa que los percentiles permiten hacer comparaciones entre sujetos en una misma prueba, y también permiten conocer el comportamiento de un

mismo individuo en diferentes pruebas. Ellos dan una idea rápida e intuitiva de la posición relativa de la persona en el grupo, al informar el porcentaje de personas a quien se iguala o supera; son de aplicación universal, ya que pueden usarse lo mismo con adultos que con niños, resultando adecuados para cualquier tipo de prueba, sea que mida variables aptitudinales o de personalidad.

Los percentiles han sido clasificados como *puntos percentiles*, en cuyo caso se divide la distribución en 100 partes iguales; *deciles*, que dividen la distribución en diez segmentos iguales (percentiles 10, 20, ...90); y *cuartiles*, que dividen la distribución en 4 partes iguales (percentiles 25, 50 y 75). En una distribución normal, el percentil cincuenta coincide con la moda, la mediana y la media de la distribución.

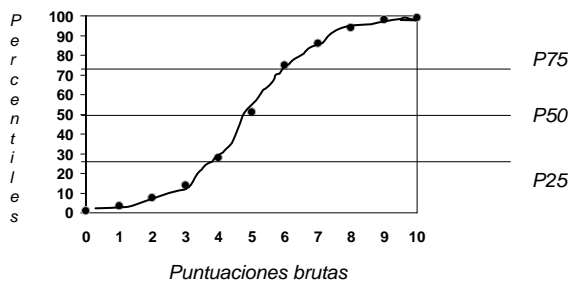


Figura 2. Calificación percentilar para cada puntuación directa

La escala percentilar cambia la forma de la distribución original dándole la forma de una ojiva, independientemente de la forma que inicialmente tuviera la distribución de los puntajes brutos; o sea, cuando se grafica la puntuación bruta con su percentil correspondiente se forma una ojiva, tal como se puede apreciar en la figura 2.

Esta alteración de la forma de la distribución original exagera las diferencias de las puntuaciones directas hacia el centro de la distribución con relación a las que se encuentran en los extremos; es decir, pequeñas diferencias de calificaciones brutas al centro de la distribución dan como resultado grandes diferencias en percentiles.

Como se puede ver en el cuadro 5, a las puntuaciones 4 y 5 le corresponden, respectivamente, el percentil 38 y 63, mientras que a las calificaciones 9 y

10 le corresponden los percentiles 99 y 100, respectivamente, de modo que siendo la diferencia en la escala de puntuaciones brutas la misma (1 punto), cuando ésta se da hacia el centro de la distribución implica 25 puntos percentilares ($63-38=25$), pero sólo 1 punto cuando se da en los extremos de la misma ($100-99=1$). A menos que se tenga esto en mente, los percentiles se pueden malinterpretar con facilidad.

Una de las debilidades de los percentiles es que al representar una *escala ordinal* permite ordenar a los sujetos, pero al no garantizar la igualdad de los intervalos no se pueden sumar, restar multiplicar ni dividir, lo cual no es importante a la hora de interpretar las calificaciones, pero sí para efectuar análisis estadísticos.

Cuadro 5

Puntuaciones originales y percentilares

Puntuaciones brutas X	f	Frecuencia acumulada fa	Percentiles P	
0	4	4	2	
1	6	10	5	
2	10	20	10	
3	16	36	18	
4	40	76	38	
5	50	126	63	
6	46	172	86	
7	10	182	91	
8	12	194	97	
9	4	198	99	<i>Puntuaciones directas</i>
10	2	200	100	$\bar{X} = 4,91$ $S = 1,86$

Para obtener los percentiles, cuando se trata de una tabla de frecuencia simple se prepara una columna de frecuencias acumuladas, el resultado se divide entre el número de personas del grupo normativo y para expresarlo en porcentajes se multiplica por 100, tal como se muestra en el cuadro 5.

Según Muñiz (1992) y Brown (1980), cuando se trate de distribución de frecuencias agrupadas se deben seguir, tal como se muestra en el cuadro 6, los siguientes pasos para elaborar el cuadro de percentiles:

1. Preparar una columna de frecuencias acumuladas (fa).
2. Determinar la frecuencia acumulada al punto medio del intervalo (fa al pm), lo cual se hace dividiendo la frecuencia simple (f) entre dos, para luego sumarle la frecuencia acumulada, así: (fa al pm): $= (f / 2) + fa$. (Ejemplo, en el intervalo 43-45 la frecuencia acumulada en su punto medio se calcula dividiendo $16/2+24=32$).
3. Dividir la frecuencia acumulada al punto medio entre el número total de participantes, para obtener la proporción de personas a ese intervalo $p = fa$ al pm / N . (Ejemplo $32/64 = .50$)
4. Multiplicar por 100 la proporción para llevarla a porcentaje de personas con una calificación dada.

Cuadro 6

Puntuaciones originales y percentilares

X	f	fa	fa al pm	$pro.$	P
46-48	24	40	52	.81	81
43-45	16	24	32	.50	50
40-42	10	14	19	.29	29
37-39	8	6	10	.15	15
34-36	6	0	3	.04	4
	64				

A una puntuación original inferior a cualquiera de las obtenidas por la muestra de tipificación le corresponde el percentil cero y a una puntuación original superior a la obtenida por la muestra de tipificación le corresponde el percentil 100; no obstante, esos percentiles no suponen una puntuación bruta de cero o perfecta.

En el ejemplo presentado en el cuadro 5, por tratarse de una distribución de frecuencia simple se puede conocer, sólo observando en la tabla de norma, el percentil que le corresponde a cada calificación bruta; así, por ejemplo, a la calificación 3 le corresponde el percentil 18 y al puntaje 8 le corresponde el percentil 99.

Sin embargo, en distribuciones de frecuencia análogas a la presentada en el cuadro 6, que se dan bajo la forma de intervalos de clases, no se puede conocer directamente el percentil correspondiente al puntaje con sólo observar el cuadro; o sea, no se tiene directamente el percentil que le corresponde al puntaje 38, por ejemplo. Con tal objetivo se debe recurrirse al conocido procedimiento de interpolación, para el cual se requiere:

1. Determinar el percentil correspondiente a cada intervalo, como se mostró antes (cuadro 6).
2. Ubicar el intervalo en el que se localiza el puntaje directo de interés (38) y los percentiles correspondiente a esos límites.
3. Interpolarse en los siguientes términos:

Puntajes	Percentiles
$2 \left[\begin{array}{c} 39 \\ 38 \\ 37 \end{array} \right] 1$	$d \left[\begin{array}{c} 15 \\ ? \\ 4 \end{array} \right] 11$
$\begin{array}{l} 2 \text{ --- } 11 \\ 1 \text{ --- } d \end{array}$	$d = \frac{11 * 1}{2} \Rightarrow d = 5,5$
$? = 4 + 5,5 \Rightarrow ? = 9,5 \approx 10$	

De tal manera que a pesar de que la distribución de frecuencias es agrupada, es posible saber que al puntaje directo 38 le corresponde el percentil 10.

Igualmente se puede determinar la puntuación que le corresponde a un percentil determinado, mediante la siguiente expresión:

$$Per = L_{inf\ erior} + \frac{(N)(p_{en.ese.intervalo} - f_{acumulada.anterior})}{f_{acumulada.del.int.ervalo}}(i)$$

donde i es la marca de clase.

Así, si se quiere conocer la calificación que le corresponde al percentil 10, se aplica la ecuación:

$Percentil(10) = 36,5 + \frac{(64)(0,15-3)}{10} * 1 = 37$ y se tiene que le corresponde la puntuación 37.

PUNTUACIONES TÍPICAS O ESTÁNDARES NORMALIZADAS

Cuando se trata de comparar puntajes de una misma persona en dos tests diferentes es preciso examinar primero la forma de la distribución de los puntajes de cada uno de dichos tests. Si las formas de las distribuciones originales son diferentes, los puntajes aun transformados a una misma escala no serán equiparables.

Esto se puede entender mejor mediante el siguiente ejemplo. Suponga que se quiere comparar la ejecución de una persona en dos escalas: una tiene una distribución normal y la otra una distribución asimétrica negativa; en la primera escala si la persona obtuvo una calificación promedio recibirá una puntuación típica de 0, una Z de 50, el percentil 50, pero si obtuvo la calificación promedio en la distribución asimétrica, aunque le corresponde una puntuación típica de 0 y una Z de 50, el percentil equivalente sería 80 y no 50 como en el caso anterior, luego las calificaciones estandarizadas no serían equiparables.

En casos como éstos y en todos aquellos en los que la forma de la distribución de puntajes directos sea no normal debido a problemas de muestreo, la solución metodológicamente pertinente es normalizar la distribución asimétrica de puntajes brutos, es decir, modificar la forma de dicha distribución original, convirtiéndola en una distribución normal, en cuyo caso se habla de *puntuaciones estándar normalizadas*. Ellas permitirán comparar las calificaciones bajo igualdad de condiciones.

Antes de efectuar transformaciones a puntuaciones normalizadas hay que tener bases para suponer que el atributo medido se distribuye normalmente en la población y adicionalmente verificar mediante una prueba de significación estadística, que es justo asumir que las puntuaciones serían normales de no ser por errores de muestreo; en caso contrario, se estarían falseando los datos al asumir como normal, una distribución que de hecho *no* se da empíricamente.

En las transformaciones no lineales se alteran las distancias relativas entre las puntuaciones de los individuos con respecto a la distribución de puntajes originales, por lo que la ecuación de regresión lineal para predecir los puntajes normalizados a partir de los originales no se puede emplear.

En lugar de ello, las puntuaciones típicas normalizadas (cuadro 7) se obtienen a través de los siguientes pasos:

1. Se determinan los percentiles correspondientes a las puntuaciones brutas, mediante el método tradicional.

Cuadro 7

Puntuaciones originales y normalizadas directas

Puntuaciones brutas X	f	Frecuencia acumulada fa
0	2	-2,06
1	5	-1,65
2	10	-1,28
3	18	-0,91
4	38	-0,31
5	63	0,33
6	86	1,08
7	91	1,34
8	97	1,88
9	99	2,4
10	100	3

2. Los percentiles se transforman en proporción, por ejemplo, el percentil correspondiente a puntaje bruto 3 es 18, por tanto, $18:100 = 0,18$; el área correspondiente se ubica en la tabla de áreas bajo la curva normal. Dado que dicha tabla suministra información del centro hacia fuera, a 0,5 que es el área total de un lado de la curva, se le resta a 0,18 lo que da un área de 0,32 que se ubica en la tabla obteniéndose una puntuación normalizada de 0,91 que es negativa, puesto que la calificación 3 está por debajo de la media de la distribución original que es de 4,91.

El cuadro 8, que muestra las puntuaciones típicas directas lineales y normalizadas, permite observar que existen diferencias numéricas entre ellas, lo cual implica que la forma de la distribución original no presentaba una forma normal.

Cuadro 8

Comparación de puntuaciones normalizadas y lineales

Puntuaciones brutas X	P	Z_n	Z_x
0	2	-2,06	-2,6
1	5	-1,65	-2,1
2	10	-1,28	-1,6
3	18	-0,91	-1
4	38	-0,31	-0,5
5	63	0,33	0,1
6	86	1,08	0,6
7	91	1,34	1,1
8	97	1,88	1,7
9	99	2,4	2,2
10	100	3	2,7

Las puntuaciones normalizadas informan a qué distancia, en términos de desviaciones típicas, se encuentra una persona con respecto a la media de la distribución, a la que se ha obligado a comportarse como una curva normal. La media de esta distribución transformada es cero y su desviación típica es igual a uno.

Una vez que la transformación a puntuaciones normalizadas se ha realizado, la interpretación de los datos es exactamente la misma que cuando se transforma a una escala típica lineal; es decir, una calificación de -1 significa que sobrepasa, aproximadamente, al 16% del grupo y una calificación de $+1$, que aventaja al 84%. Igualmente, presenta sus mismas limitaciones: el usar números muy pequeños y negativos, por lo cual es frecuente recurrir a las puntuaciones típicas normalizadas derivadas que se detallan seguidamente.

PUNTUACIONES TÍPICAS O ESTÁNDARES NORMALIZADAS DERIVADAS

Dado que al relacionar las puntuaciones originales x con las típicas normalizadas, se observa una relación lineal y que la correlación entre tales calificaciones es positiva perfecta, se puede proceder del mismo modo que en las transformaciones típicas lineales derivadas, en el sentido de recurrir a la ecuación de regresión lineal simple para llevar las calificaciones típicas normalizadas a otras escalas, que al poseer valores de media y desviación numéricamente más grandes, permitan efectuar análisis más sencillos una vez superadas las limitaciones de emplear números negativos o con decimales. Tal ecuación, como se señaló antes, se expresa como

$$Y' = \frac{sy}{sx}(x - \bar{x}) + \bar{y}$$

pero dado que la transformación no implica cambio en

la posición relativa de los sujetos, la correlación entre los puntajes originales y los transformados será de uno, por lo que la ecuación se puede escribir como $Y' = \bar{y} + (z_x)(s_y)$.

Igual que la calificación estándar (hasta aquí has puesto estándar sin plural) derivada linealmente, la normalizada también pueden asumir cualquier forma conveniente.

Una escala normalizada derivada que se usa con cierta regularidad dentro de la comunidad científica es la conocida como T , propuesta por McCall en 1922, la cual tiene media 50 y desviación 10. La transformación a la escala normalizada derivada t se obtiene al aplicar la ecuación anteriormente referida, de modo que si por ejemplo se debe transformar el puntaje bruto 9 a T se determina su z_n y de seguida la ecuación $Y = 50 + (2,4 * 10) = 74$, con lo cual se desprende una puntuación de 74. Lo anterior implica que una persona que obtiene un puntaje bruto de 9 le corresponde una T de 74 y se encuentra a dos desviaciones y media por encima de la media de la distribución transformada.

Otro tipo de puntuación normalizada derivada es la *estanina o eneatipo*, cuyo nombre es un derivado de *standard nine*, pues se divide el rango total de las puntuaciones en 9 intervalos y se asigna a cada uno un número del 1 al 9. Ésta representa una escala derivada de media 5 y desviación típica 2.

Se dispone de diversos procedimientos para transformar una puntuación directa a estanina: o través de la ecuación de la recta o mediante la tabla de percentiles.

Cuando se aplica la ecuación de la recta se usan las constantes propias de esta escala, a saber, media 5 y desviación 2; así $E = 5 + z_n * 2$ para lo cual, como se puede apreciar, se recurre al uso de puntuaciones normalizadas directas z_n . Así, a una persona quien calificó con 4 le corresponde casualmente eneatipo 4, ya que $E_4 = 5 + (-2,8 * 2) = 4$.

Cuadro 9

Puntuaciones y normalizadas derivadas T y E

<i>X</i>	<i>Z_n</i>	<i>T</i>	<i>E</i>	
0	-2,33	27	1	
1	-1,81	32	2	
2	-1,44	36	2	
3	-1,08	39	3	<i>Puntuaciones</i>
4	-0,58	44	4	<i>directas</i>
5	0,01	50	5	$\bar{X} = 4,91$
6	0,66	57	6	$S = 1,86$
7	1,20	62	7	<i>Escala T</i>
8	1,56	66	8	$\bar{X} = 50$
9	2,05	71	9	$S = 10$
10	2,57	75	9	<i>Escala E</i>
				$\bar{X} = 5$
				$S = 2$

En el cuadro 9 es posible observar las puntuaciones directas y sus correspondientes transformadas a la escala *T* que, como se dijo, tiene media 50 y desviación 10, así como las puntuaciones *E*, cuya media, como se recordará, es de 5 y su desviación es de 2; esta última se registra calculada por medio de la ecuación de la recta.

Un procedimiento alternativo para determinar la estanina, es la tabla de percentiles reportada en los apéndices de los textos de psicometría, análogas a la figura 3. Si, por ejemplo, se quiere determinar la estanina del puntaje bruto 8 del ejemplo que se viene trabajando, se debe: 1) ubicar qué percentil le corresponde y, como se puede ver en el cuadro 5, el del puntaje 8 es el percentil 97; 2) se localiza tal percentil en la figura 3, donde se puede apreciar que le corresponde el eneatipo 8. Luego, se puede afirmar que una persona quien calificó con 8 puntos se ubica en la estatina 9.

Como se observa, hay disenso entre las estaninas determinadas por la ecuación de regresión y por la tabla de áreas bajo la curva normal, prefiriéndose siempre esta última opción que resulta la más precisa.

Distribución normal

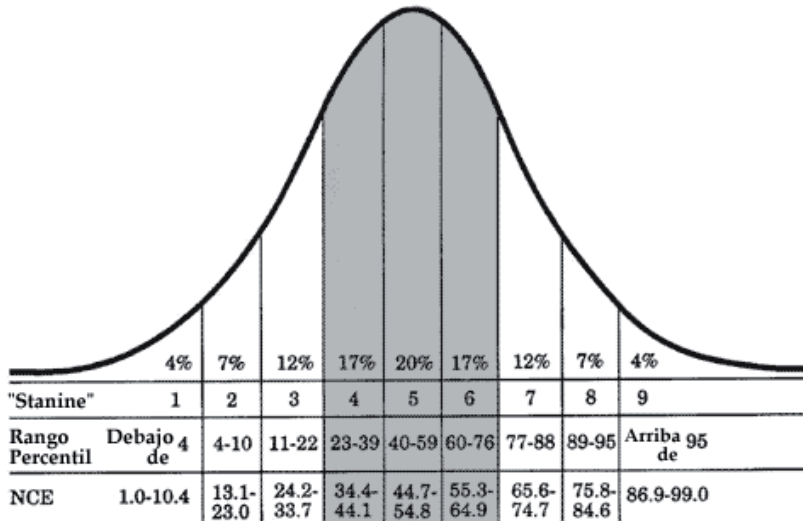


Figura 3. Calificación percentilar para cada puntuación original

Los eneatis representan una escala poco recomendable, ya que se pierde mucha información al incluir en el mismo eneatis a personas con muy distintas puntuaciones; no obstante, tiene la ventaja de ser de fácil interpretación

Por su lado, el otro tipo de puntuación normalizada derivada como lo son las puntuaciones *T*, permiten una diferenciación más fina entre los puntajes que las estaninas. Así, para una confiabilidad de 0,91 el error estándar de medida ($s_m = s_r \sqrt{1-r_{tt}}$) será en una escala *T* de 3 puntos, pero en una escala de estanina será de 0,6 puntos; ello significa que en esta última escala cualquier variación de centésimas es vital. La escala *T* ofrece como ventaja comparativa el tener unidades más amplias que facilitan la diferenciación.

EVALUACIÓN DE LAS PUNTUACIONES TRANSFORMADAS

Puesto que las calificaciones típicas ya sean lineales o normalizadas expresan las puntuaciones de una prueba en una escala de intervalos, resultan especialmente útiles cuando se necesitan análisis estadísticos posteriores, como el combinar calificaciones en una medida compuesta.

Por otro lado, las calificaciones estándares permiten hacer comparaciones directas entre los puntajes de una persona en dos o más pruebas, ya que una calificación típica representa la misma posición relativa en las dos distribuciones independientemente de la media y la desviación de la distribución original.

Cuando la distribución de calificaciones directas se aproxima a la curva normal, las puntuaciones típicas lineales y normalizadas son muy similares, a pesar de las diferencias en los procedimientos empleados para derivarlas. De hecho, normalizar una distribución que ya lo es producirá poco o ningún cambio.

Siempre será preferible hacer que una distribución sea normal a partir de manipulaciones a nivel de la dificultad de los ítems, reservándose la normalización a través de puntuaciones típicas, sólo para cuando se sospeche que la asimetría de la distribución original es producida por errores de muestreo.

Las calificaciones estándar pueden transformarse directamente a una escala de percentiles que es de más fácil interpretación para la mayoría.

ESCALAS DE DESARROLLO

Muchas capacidades, habilidades y características se desarrollan sistemáticamente a través del tiempo; así, por ejemplo, la capacidad de lectura de un niño o su inteligencia se desarrollan con la edad, los años de asistencia a la escuela y las interacciones con su medio ambiente. Cuando interesa desarrollar un instrumento para examinar una capacidad que muestra las características referidas, esto es, que aumenta sistemáticamente con el paso del tiempo, se pueden elaborar escalas que comparen la ejecución de un individuo con la persona promedio a varios niveles desarrollo de la variable, denominadas en la literatura con el nombre genérico de escalas de desarrollo.

Según Anastasi y Urbina (1996), hay cuatro tipos de escalas de desarrollo: la edad mental, el grado escolar, las escalas ordinales y los cocientes, cada uno de ellos poseen características particulares que los hacen útiles bajo determinadas circunstancias. Este tipo de escalas resulta apropiado cuando la característica medida cambia sistemáticamente con la edad, por lo que son usadas regularmente en poblaciones infantiles; en cualquier caso su uso exige gran pericia, pues como se verá seguidamente, las unidades en que ellas se expresan no son iguales en los diversos niveles de desarrollo.

ESCALAS DE EDAD MENTAL

Fueron ideadas por Binet a finales del siglo XIX, para medir el crecimiento mental de los niños en función de su desempeño, con reactivos en los que la ejecución variaba con la edad. En la actualidad son empleadas para medir cualquier variable que cambia con el paso de los años.

El elaborador de un instrumento expresado en una escala de edad mental, administra el test a participantes de un mismo grupo normativo de diferentes edades y determina el número de ítems que la mayoría de los niños de cada edad contesta correctamente.

Durante la etapa de usuario, el examinador le asignará a cada evaluado la edad mental que le corresponda, según el número de ítems contestado correctamente.

Por ejemplo, en un test que mide una variable que se desarrolla con la edad, se encontraron los resultados mostrados en el cuadro 10, donde se observa, por ejemplo, que la mayoría de los niños de 1 año contestó correctamente entre 0 y 4 reactivos y que la mayoría de los niños de 3 años contestó correctamente entre 9 y 12 reactivos. Cuando el instrumento se administre a una persona, independientemente de su edad cronológica, y ella conteste correctamente entre 13 y 17 reactivos, se dirá que le corresponde una edad mental de 4 años en este instrumento.

Cuadro 10
Escala de edad mental

<i>Edad mental</i>	<i>Ítems correctos</i>
1	0-4
2	5-8
3	9-12
4	13-17

Los niños cuya edad cronológica es menor (2 años) que su edad mental (4 años) habrían alcanzado un nivel de desarrollo intelectual más alto que el de la mayoría del grupo de su misma edad cronológica; lo contrario podría decirse de un niño cuya edad cronológica sea más alta que su edad mental.

Las escalas de edad mental pueden proveer la información semanal, mensual, trimestral, semestral o anualmente, ello dependiendo de la rapidez con que se desarrolle en el tiempo la variable medida y la precisión que desee mostrar quien construye la prueba.

Si bien este tipo de escalas son útiles para medir atributos que crecen o decrecen con la edad, ellas no poseen intervalos iguales. Así, la mayoría de las capacidades intelectuales se desarrollan con rapidez y a un índice bastante constante durante toda la infancia, pero comienza a desacelerarse en la adolescencia; consecuentemente, el desarrollo de la capacidad de lectura, por ejemplo, será menor entre los 17 y 18 años que entre los 7 y 8 años. De modo que la misma diferencia (1 año) tendrá diferentes implicaciones a diferentes edades.

La salida a la limitación citada, es elaborar y emplear antes este tipo de escalas para medir variables que se desarrollen con la edad y sólo dentro de los límites en que dicho desarrollo ocurra. Así, no debe emplearse una escala de edad para medir inteligencia a los 18 años, porque es sabido que a esa edad y en adelante la variable se ha estabilizado.

Otro aspecto que se debe tomar en consideración a la hora de elaborar escalas de edad mental, es la necesidad de contemplar en el grupo de referencia a sujetos con características comunes sometidos a circunstancias más o menos similares, tanto de educación como de experiencias, ya que ellas pueden afectar el nivel de desarrollo alcanzado y distorsionar la escala.

ESCALAS DE GRADOS ESCOLARES

Representan un tipo análogo a las calificaciones de edad mental, pero comparan la ejecución de una persona con la del promedio de los estudiantes de varios grados. De modo que si el grado equivalente de un niño es de 3 grados-5 meses, significará que su ejecución será comparable al alumno promedio del quinto mes de tercer grado.

Las calificaciones de grados escolares son útiles si se presumen experiencias educativas similares, en caso contrario, las puntuaciones no son necesariamente equivalentes. Así, si en una prueba de matemáticas de 75 preguntas (50 de aritmética y 25 de álgebra) la calificación promedio para alumnos de 8^{vo} grado es 40 y de 9^{no} 50, siendo que sólo se dicta álgebra en 9^{no}; la misma calificación de 50 en un alumno de 8^{vo} grado significará cosas diferentes que para un alumno de 9^{no}. Mientras que para el alumno de 8^{vo} tiene un excelente manejo de aritmética (hizo bien todas las preguntas de su grado) el de 9^{no} maneja algunas cosas de aritméticas y algunas de álgebra.

Cuando se elaboran escalas de grados escolares, se debe tener especial cuidado con la conformación del grupo normativo, de modo de incluir en cada grado sujetos típicos y no alumnos avanzados o pocos aventajados, ya que ello distorsionaría la escala. En tal sentido, Brown (1980) recomienda el uso de normas modales de edad, las cuales consisten en incluir dentro del grupo normativo sólo estudiantes cuya edad cronológica se corresponda con el grado cursado.

ESCALAS ORDINALES

Las escalas ordinales son una modalidad de las escalas de desarrollo que permiten el ordenamiento de los sujetos sin importar la diferencia numérica entre ellos. Se emplean principalmente en funciones como la locomoción, la discriminación sensoriomotora, la comunicación lingüística y la formación de conceptos.

Una escala ordinal para un área, consiste en un protocolo, en el que para cada edad (semanas o meses) se describen las conductas características después de una observación detallada de muchos individuos. Se dice que es ordinal, ya que se sigue un orden constante en el que cada etapa (edad) presupone el dominio de las conductas características de etapas previas.

Cuando se aplica una escala ordinal, el evaluador debe precisar las conductas que emite el niño evaluado y comparar esa ejecución contra el protocolo. La calificación obtenida por el niño será la edad que corresponda a la que es típica para su ejecución.

Si bien los resultados de una escala ordinal se expresan en términos de niveles aproximados de edad, estos números son secundarios y lo relevante es la descripción cualitativa del comportamiento característico del niño que informa cuáles conductas posee en su repertorio y cuáles no.

Una de las escalas ordinales más famosas es la presentada por Gesell y sus colegas en 1947. En ella el nivel de desarrollo de un niño se encuentra comparando su conducta, con la que es distintiva de ocho edades clave que van desde las 4 semanas a los 6 meses.

COCIENTES Y RAZONES

La edad mental y los grados escolares no informan acerca del nivel de desarrollo relativo. Un mismo puntaje, por ejemplo, 10, tiene diferentes significados si pertenece a una persona de 8, 10 o 15 años, no obstante dichas escalas no ofrecen ninguna información al respecto. En tal sentido, se ha fomentado el uso de las escalas de cociente intelectual y educativo.

COCIENTE INTELECTUAL

Wilhelm Stern propuso relacionar la edad mental con la edad cronológica, como un índice del ritmo de desarrollo, al que se denomina coeficiente o cociente intelectual. De modo que

$$CI = \frac{\text{Edad . Mental}}{\text{Edad . Cronológica}} 100$$

De la ecuación anterior pueden derivarse ciertos aspectos importantes con relación al coeficiente de inteligencia:

a) El *CI* será de 100 para todos los niños que tengan el mismo nivel de inteligencia que el niño promedio de la misma edad, éstos son los que tienen la misma edad mental y cronológica.

b) Los que tengan un nivel de desarrollo intelectual más bajo que el promedio de su grupo (edad mental menor que la cronológica), tendrá un *CI* menor de 100.

c) El *CI* será mayor que 100 en los niños que se hayan desarrollado más rápidamente que sus compañeros (edad mental mayor que la cronológica).

Para hacer las interpretaciones expuestas antes, es necesario que la desviación típica de la distribución de los coeficientes intelectuales sea más o menos la misma para todos los grupos de edades.

Debido a que dentro de un mismo test las desviaciones estándares no son exactamente la misma en las diferentes edades, un mismo *CI* no da la misma posición relativa en las distribuciones de las diferentes edades. En vista de esta debilidad, las razones de *CI* ya no se utilizan en las principales pruebas de inteligencia y en su lugar se emplean puntuaciones estándar normalizadas derivadas, basadas en una muestra representativa de la población de cada nivel. Tales calificaciones denominadas *CI* de desviación (*CI*s), tiene una media de 100 y desviación típica de 15 (escalas de Wechsler) o 16 (Stanford-Binet) en cada nivel, de edad y se calculan en los mismos términos que se explicó en este documento, en la sección de puntuaciones normalizadas derivadas.

El vocablo *CI* utilizado con frecuencia en el mundo de la psicología actual, es un sustituto del vocablo correcto *CI* de desviación, éste lleva implícito una cantidad de elementos técnicos que deben ser mantenidos en mente. Por un lado, recoge el concepto de edad mental, pero lo dimensiona al darle la noción de relatividad comparándolo con la edad cronológica; y por otra parte, le agrega la cualidad de puntuación normalizada, lo que implica que la distribución original se normaliza para que sus unidades puedan ser comparables. Todas estas fortalezas explican por qué el *CI* de desviación se emplea casi universalmente, cada vez que se preparan escalas de desarrollo.

COCIENTE EDUCATIVO Y DE RENDIMIENTO

Tienen como finalidad comparar las realizaciones verdaderas de una persona con su ejecución esperada. Para su cálculo se utilizan como numerador alguna medida de ejecución y como denominador la edad cronológica, alguna medida de capacidad intelectual o la ubicación en un grado escolar.

Existen dos desventajas, por un lado, la razón entre dos calificaciones no confiables será menos confiable aún; y por otro, compara una medida de rendimiento con otra de capacidad intelectual, con lo cual se puede afirmar que parten del supuesto de que el rendimiento está determinado exclusivamente por la capacidad intelectual, y esto no es consistente con los hechos empíricos. No obstante, las limitaciones de los cocientes escolares, algunas pruebas aún lo utilizan. Por esta razón se consideró referirlos en este texto y puntualizar sus debilidades que deberán tomarse en cuenta si requiere emplearlos.

TIPOS DE NORMAS Y GRUPO DE REFERENCIA

Las calificaciones normativas se pueden preparar en función de todos los habitantes de un país o de grupos más restringidos. En atención a ello se habla de *normas nacionales* y *normas específicas*.

NORMAS NACIONALES

Las normas nacionales son elaboradas sobre una muestra de tipificación representativa de la población de todo un país. Para definir tales normas se seleccionan los miembros de la muestra, sobre la base de un muestreo proporcional y estratificado, que es el que garantiza representatividad de poblaciones heterogéneas. Este tipo de normas es poco usual, no sólo por lo costosa en términos de tiempo y recursos, sino porque es tan general que proporciona poca información en torno a grupos particulares.

NORMAS ESPECÍFICAS

Se elaboran al estandarizar la prueba con poblaciones restringidas y escogidas para adaptarse a sus propósitos particulares. Este tipo de normas es la más apropiada para cualquier instrumento y su generalización viene dada por la definición de la población.

Resultan útiles cuando se identifican puntuaciones apreciablemente diferentes para distintos grupos, de modo que se conforman subgrupos según género, edad, grado escolar, programa de estudio, región geográfica, ambiente urbano o rural, nivel socioeconómico, entre otras variables. El uso que se le vaya a dar a la prueba es el que determina cuál de esas variables es la más relevante.

Dentro de las normas específicas cabe nombrar las normas locales. Éstas son establecidas por el propio usuario del instrumento en sus ambientes particulares, cuando se da cuenta de que ninguno de los datos normativos disponibles en el manual se adapta a sus fines. La muestra empleada se define de un modo más limitado que los subgrupos de normas específicas, en tanto que las calificaciones se expresan como una distribución de frecuencia y eventualmente se presentan como puntuaciones transformadas.

Estas normas locales son más apropiadas que las nacionales para muchos propósitos de evaluación, ya que cada persona es comparada con sus asociados inmediatos, cuya ejecución puede ser diferente de los del grupo normativo nacional.

La mayor debilidad de las normas locales es que no permiten interpretaciones en contextos más amplios que los provistos por la muestra de normalización. El punto de quiebre está en manos del examinador, quien decidirá lo más conveniente para su caso particular y actuará en consecuencia.

MÉTODOS PARA PRESENTAR DATOS NORMATIVOS

Existen distintas técnicas para resumir y presentar datos normativos, éstas se pueden ubicar en uno de los tres grupos siguientes: *los cuadros de conversión simple y múltiple*, y *los perfiles*.

CUADROS DE CONVERSIÓN SIMPLE

Los cuadros de conversión simple muestran las calificaciones brutas y las transformadas equivalentes, ya sean en calificaciones estándar, en percentiles o en escalas de desarrollo, para un grupo normativo específico en una sola prueba.

Cuadro 11
Cuadro de conversión simple de la PAU

<i>Puntajes brutos</i>	<i>P</i>	<i>Z</i>
19	99	70
18	85	65
17	72	60
16	60	54
15	50	50
14	44	42
13	32	25
12	20	30
11	4	26

El cuadro 11 es de conversión simple que muestra, para cada calificación bruta, los rangos percentilares y, calificaciones típicas lineales derivadas de un grupo de personas en una prueba aptitud universitaria (PAU). Mediante dicho cuadro es posible ubicar de manera inmediata que a la persona con una calificación bruta de 19 le corresponde el percentil 99 y una puntuación típica lineal derivada de 70 puntos, lo cual implica que está a dos desviaciones típicas por encima de la media.

CUADROS DE CONVERSIÓN MÚLTIPLE

Suponen presentar en un mismo cuadro de conversión datos de varias pruebas distintas, subpruebas de una batería o normas locales de una misma prueba.

Cuando se trate de pruebas distintas o subpruebas de una batería es indispensable utilizar el mismo grupo normativo para todas ellas; es este caso el cuadro de conversión múltiple proporcionará un método para comparar simultáneamente la ejecución de una misma persona en dichas pruebas o subpruebas.

Cuadro 12

Cuadro de conversión múltiple de una batería

<i>Puntajes brutos</i>	<i>Verbal</i>	<i>Numérico</i>	<i>Abstracto</i>
19	99		
18	85		99
17	72	99	80
16	60	87	73
15	50	72	60
14	44	60	40
13	32	45	25
12	15	30	10
11	1	15	1
10		9	
9		1	

En el cuadro 12 se puede comparar el desempeño de una misma persona en tres pruebas diferentes (verbal, numérica y abstracta), al transformar su calificación bruta en percentiles, partiendo de que la muestra normativa es la misma.

Es posible también preparar un *cuadro de conversión múltiple* para mostrar las normas específicas de una misma prueba, quedando de parte del usuario ubicar aquel grupo particular dentro del cuadro de conversión múltiple que mejor se adapte a sus fines.

PERFILES

Los perfiles son representaciones gráficas que permiten visualizar rápidamente el desempeño de un examinado. Pueden ser diseñados de modo que muestren de una sola ojeada la configuración de las calificaciones de una persona en diversas pruebas, o que permitan comparar las puntuaciones de varios examinados en una misma prueba.

Es indispensable utilizar para todas las pruebas involucradas el mismo grupo normativo, en caso contrario, las calificaciones no serán directamente comparables. Adicionalmente, las puntuaciones se deben expresar en la

misma escala, por lo cual se tiene la necesidad de utilizar dimensiones que no exageren las diferencias pequeñas o reduzcan las grandes. Una vía para ello es presentar las calificaciones en bandas en lugar de hacerlo como puntuaciones exactas, de modo que las diferencias entre ellas sólo sean significativas si las bandas no se superponen.

Z	Verbal P. bruto	Numérico P. bruto	Abstracto P. bruto	Z
90				90
80	14-15		61-70	80
70	12-13	81-100	51-60	70
60	10-11	61-80	41-50	60
50	8-9	41-60	31-40	50
40	6-7	21-40	21-30	40
30	4-5	0-20	11-20	30
20	2-3		0-10	20
10				10

Figura 4. Perfil de una batería de aptitudes

A través de perfiles como el mostrado en la figura 4 se puede no sólo transformar las calificaciones originales a una misma escala (calificación estándar lineal), sino que además se observa gráficamente el comportamiento de los examinados. Para la persona del ejemplo se aprecia que su mejor ejecución fue en la subprueba verbal ($Z=80$), pese a que su puntaje bruto fue el más bajo de todos; el peor desempeño lo obtuvo en la subescala numérica ($Z=40$), donde se ubica a una desviación típica por debajo de la media de la escala Z.

EQUIPARACIÓN DE PUNTUACIONES

Cuando se emplean normas es posible comparar una persona con su grupo normativo o con un sujeto particular del mismo grupo normativo y también es posible comparar a una persona consigo misma en diferentes instrumentos; en estos casos es importante recurrir a lo que Muñiz (1992) ha dado en llamar equiparabilidad de puntuaciones.

La equivalencia o equiparación de puntuaciones es el proceso de desarrollar una conversión del sistema de unidades de un test al sistema de unidades del otro, de modo que las puntuaciones derivadas de ambos después de la conversión sean intercambiables.

Para hablar apropiadamente de equiparabilidad entre puntuaciones de dos tests, éstos deben estar basados en una misma población normativa, poseer las similares propiedades psicométricas como confiabilidad, validez y grado de dificultad y tener la misma finalidad.

Se puede hablar de tres diseños utilizados clásicamente con estos fines: un solo grupo, grupos equivalentes y test de anclaje y dentro de cada uno de ellos se puede recurrir a distintos métodos: media, transformación lineal y percentiles. A continuación se presentan los diseños y luego los métodos.

DISEÑOS PARA LA EQUIPARACIÓN

Un solo grupo

Consiste en elegir una muestra aleatoria de personas a los que se les aplica ambos tests y a través de la media, la transformación lineal o los percentiles se procede a la equiparación. Regularmente se usa el contrabalanceo para corregir los efectos del orden.

Dos grupos equivalentes

Se selecciona dos muestras aleatorias de la población a la que se destinan los dos tests y se le aplica uno de los tests a cada una de ellas, procediéndose luego a la equiparación bajo el supuesto de que el azar generó muestras equivalentes de sujetos.

Test de anclaje

Se aplican los dos tests que se equiparán a dos muestra (uno a cada una), y se aplica cierto número comunes de ítems de anclaje que permitan establecer las equivalencias entre los tests a equiparar. Aquí las dos muestras no tienen necesariamente que ser equivalentes.

MÉTODOS PARA LA EQUIPARACIÓN

Entre los métodos para equiparar las puntuaciones de un par de tests se puede recurrir a la media. Suponga que para el test X la media es $x = 40$ y

que para el test Y la media es $y = 45$; la equiparación se establece sumando 5 puntos a las calificaciones de un sujeto en el test X para equipararlas a Y . Si la conversión va a la inversa se restan 5 puntos. Esta transformación se usa poco.

Otro método de equiparación es a través de una conversión mediante la *ecuación de regresión lineal*, tal como la que se expuso cuando se discutieron las puntuaciones típicas (lineales o normalizadas) derivadas. Así, se transforma la calificación de una persona a puntuaciones Z_x y posteriormente se hace una transformación derivada usando la media y la desviación típica de la escala a la cual se quiere hacer la conversión. Ahora se tienen las dos puntuaciones en la misma escala y la equiparación es posible.

Finalmente, pueden usarse *los percentiles* que consiste en equiparar aquellas puntuaciones de tests cuyos percentiles son iguales. Por ejemplo, si en un test X a una puntuación directa de 10 le corresponde el percentil 80 y en otro test Y a una puntuación 25 le corresponde el mismo percentil 80, la puntuación 10 en el test X es equivalente a 25 en el test Y . Así que obtener 10 en X equivale a obtener 25 en Y . Las ventajas y limitaciones de este método son las mismas que las inherentes a los percentiles.

El diseño de equiparación más común consiste en tener un solo grupo y recurrir a la conversión mediante la ecuación de regresión lineal, prestando extrema atención al grupo normativo.

RELATIVIDAD DE LAS NORMAS

Las normas no son absolutas, universales o permanentes. Ellas están sujetas al propósito de la prueba, a la población a la cual se dirigen y a la muestra de estandarización empleada para estimarlas.

La posición relativa de una persona puede variar según el *grupo normativo* con el cual se compare, motivo por el cual su composición es un factor crucial. Para conformarlo se debe tener claro en qué *poblaciones* y con qué *finés* se utilizará la prueba. Así, si será empleada para examinar la personalidad de los adolescentes, el grupo normativo lo representará un corte transversal de los adolescentes, pero si se quiere evaluar la disposición a la lectura de niños de preescolar, el grupo normativo lo constituirán preescolares que no hayan recibido instrucción para la lectura.

Lo ideal es que la población se defina de antemano según los objetivos de la prueba, pero razones prácticas pueden hacer que ello resulte inalcanzable, en cuyo caso es mejor definir la población de modo más reducido antes que aludir normas de una población ideal que no está bien representada en la muestra de tipificación. Se deberá restringir esa población de manera que se ajuste a la muestra disponible.

Para garantizar que el grupo normativo sea representativo de la población se deberá recurrir a *procedimientos de muestreos probabilísticos*, ya sea al azar simple, sistemático o proporcional y estratificado, según sean las propiedades de la población y los recursos disponibles. El autor del instrumento deberá describir en el manual de prueba, el procedimiento de muestreo empleado para garantizar tal representatividad, e incluir información demográfica como género, edad, educación, nivel socioeconómico, región geográfica, raza, así como otros aspectos como cantidad de adiestramiento especial, inteligencia, educación o cualquier otra variable que resulten de interés para el constructo considerado. De este modo suministra al usuario criterios ciertos para saber si la prueba posee un grupo normativo que se adecua a las características de las personas a las que él debe evaluar.

Si el elaborador de normas identifica que varios subgrupos de la población tienen un comportamiento (media y desviación) diferente en la prueba y dichos grupos resultan de interés para los fines del instrumento, se verá en la obligación de construir normas separadas para cada subgrupo y reportarlas en el manual de prueba. Al usuario corresponde seleccionar aquellas normas que correspondan a la persona examinada.

El grupo normativo deberá ser de tamaño moderado, pues si bien el error de muestreo varía en proporción inversa al tamaño de la muestra, es preferible disponer de una muestra representativa pequeña que una grande poco definida. En tal sentido, se debe tener cuidado de no seleccionarla sólo porque es de fácil acceso como, por ejemplo, medir aspectos de la conducta infractora sólo a través de prisioneros, ya que los propios factores que determinaron tal internación puede hacer que ese grupo no sea representativo de toda la población de infractores.

Dado que la mayoría de las pruebas se diseñan para utilizarse con varios grupos y/o con diversos fines, el autor generalmente reporta normas para más de un grupo normativo, quedando de parte del usuario seleccionar el

más apropiado para él, de modo que a las calificaciones de un estudiante de bachillerato que quiera estudiar Ingeniería, pueden ser comparadas con las de otros graduados de bachillerato y con las de los estudiantes de primer semestre de Ingeniería. Si la persona evaluada es un estudiante universitario y la población contra la que se está comparando es un grupo de niños de 14 años, su puntaje no podrá considerarse especialmente sobresaliente si el test mide alguna forma de ejecución intelectual, pero si la muestra de comparación es un grupo de estudiantes universitarios será posible un análisis más preciso.

Una vez que una prueba se presenta en forma de manual es porque se han preparado las normas y los estudios de confiabilidad y validez, pero este no es un estado definitivo, sino que dicha prueba permanece en una *etapa de revisiones* en la que elaborador y usuario se compromete a darse *feedback* para mantener las normas actualizadas. Resulta técnicamente inapropiado y antiético emplear pruebas con normas obsoletas o desactualizadas.

Antes de concluir se debe señalar que los datos normativos no son siempre necesarios. Por ejemplo, si se va a hacer una selección simple, se aceptará a los individuos que obtengan los puntajes más altos; tampoco son necesarias las normas en pruebas enmarcadas en el campo de la investigación o en pruebas criterioles que, como se sabe, recurren a tipos de calificaciones particulares.

REFERENCIAS

- Anastasi, A. y Urbina, S. (1998). *Test psicológicos*. (7^{ma} ed.). México: Prentice Hall.
- Brown, F. (1980) *Medición en psicología y educación* (4^{ta} ed.). México: El Manual Moderno.
- Ghiselli (1964). *Theory of psychological measurement*. New York: McGraw-Hill.
- Magnusson, D. (1975). *Teoría de los tests*. México: Biblioteca Técnica de Psicología.
- Muñiz, J. (1992). *Teoría clásica de los tests*. Madrid: Pirámide.